4

LETI's SCIENTIFIC SERVICES

COLECCIÓN
LETITRAIN

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado? (2ª parte)

leti laboratorios

www.leti.com/zonaprofesional

COLECCIÓN
LETITRAIN

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(2ª parte)

José Luis Mateos, PhD

"... El curso tradicional de lectura de la literatura médica consiste en: '¡Aquí tiene The New England Journal of Medicine'; *léala!*

Este método es análogo al de aprender a nadar por el método de la inmersión total.

Desde luego, ciertas personas pueden aprender a nadar de esa forma, pero algunas se ahogan y muchas le toman miedo al agua. Leer solo los resúmenes de los artículos médicos es como tener miedo al agua..."

(Riegelman RK, Hirsch. En: *Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica.* Biblioteca de la Organización Panamericana de la Salud. Little, Brown & Co. [Eds.] 1992. Boston, MA, USA)

© 2015, Laboratorios Leti
Realización editorial, Lexic S.L.
Passeig del Ferrocarril, 339, 2º 4ª - 08860 Castelldefels (Barcelona), España.

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(2ª parte)

ÍNDICE

Índice

Introducción 1

1. Criterios específicos para la lectura crítica de artículos 5

1.1. Tipos de estudios en investigación clínica 10

1.1.1. Estudios transversales 10

1.1.2. Estudios de casos y controles 13

1.1.3. Estudios de cohortes 17

1.1.4. Ensayos clínicos aleatorizados 21

1.1.5. Revisiones sistemáticas y metaanálisis 29

1.2. Lectura crítica de artículos con diseño transversal 35

1.3. Lectura crítica de artículos de casos y controles 38

1.4. Lectura crítica de artículos de cohortes 42

1.5. Declaración STROBE 49

1.6. Lectura crítica de artículos de estudios experimentales 52

1.6.1. Sobre la validez interna del estudio. ¿Son válidos los resultados de un ensayo clínico? 53

1.6.2. Valoración de la importancia del estudio. ¿Cuáles son los resultados? 57

1.6.3. Sobre la validez externa del estudio. ¿Son aplicables los resultados del ensayo? .. 60

1.6.4. Declaración CONSORT 64

1.7. Lectura crítica de artículos de revisión y metaanálisis 65

1.7.1. Sobre la validez interna de la revisión o metaanálisis 65

1.7.2. Relevancia de los resultados de la revisión o metaanálisis 71

1.7.3. Validez externa de la revisión o metaanálisis 73

1.7.4. Declaración QUORUM 74

2. **Bibliografía** 77

3. **Autor** 83

4. **Notas** 87

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(2ª parte)

INTRODUCCIÓN

En numerosas ocasiones los clínicos se enfrentan a situaciones en las que tienen que valorar el riesgo de aparición de efectos perjudiciales para sus pacientes como consecuencia de las intervenciones terapéuticas que recomiendan. Para realizar estos juicios los profesionales necesitan unos conocimientos mínimos que les permitan evaluar de forma crítica las evidencias científicas disponibles sobre etiología. Una adecuada valoración previa puede evitar conclusiones erróneas, tanto en el sentido de atribuir un determinado efecto a un agente terapéutico cuando en realidad no lo produce (falsos positivos), como en el sentido contrario, no atribuir el efecto a dicho agente cuando en realidad sí lo produce (falso negativo). Surge entonces el concepto de medicina basada en la evidencia. En el tercer libro de esta colección, también sobre la lectura crítica de un artículo, proponíamos la medicina basada en la evidencia como un proceso sistemático de búsqueda, evaluación y utilización de los hallazgos de la investigación biomédica como base esencial para la toma de decisiones en la práctica clínica. Es decir, ya no se trata de validar, sino de negar validez a todo lo que no demuestre valía previamente y, para ello, se establece toda una metodología de análisis de la información que procede fundamentalmente de los ensayos clínicos, pero no de cualquier ensayo clínico, sino de aquellos realizados de forma prospectiva, controlada y aleatoria.

El especialista ha de llevar una sistemática de trabajo en la que la lectura crítica de cualquier artículo científico que caiga en sus manos será uno de los pilares básicos que seguir dentro de la búsqueda de la evidencia: la lectura crítica, también conocida como evaluación crítica *(critical appraisal)*. Evaluar críticamente un artículo científico significa valorar su grado de evidencia o, lo que es lo mismo, saber valorar críticamente la validez –aproximación a la verdad– de dicho artículo, saber interpretar los resultados y poder determinar su utilidad a cada paciente –el potencial de aplicabilidad clínica–.

Los conocimientos sobre un tema o un campo de investigación clínica se adquieren básicamente a través de la literatura que se genera a partir de dicha investigación. Sin embargo, la búsqueda de bibliografía, la identificación de documentos o trabajos relevantes y la lectura prolongada y ordenada de revistas de nuestra especialidad no conducen a ninguna parte si no hay una lectura crítica de los artículos científicos.

Dirigido fundamentalmente a los profesionales sanitarios, este libro es la segunda entrega sobre el tema y en él se definen los distintos tipos de estudios que normalmente nos encontramos en la publicación de resultados en investigación clínica, así como los pasos que hay que seguir sobre la lectura crítica de cada uno de estos tipos de estudio.

Lectura crítica de un artículo científico
¿Cómo analizar un artículo publicado?
(2ª parte)

1

CRITERIOS ESPECÍFICOS PARA LA LECTURA CRÍTICA DE ARTÍCULOS

1. CRITERIOS ESPECÍFICOS PARA LA LECTURA CRÍTICA DE ARTÍCULOS

En la primera parte de este libro sobre cómo hacer la lectura crítica de un artículo científico ya se hizo una primera descripción de los principales diseños utilizados en la investigación médica. Atendiendo al criterio de estudios controlados o no controlados, en este apartado nos dedicaremos a ver el esquema de estudio de cada diseño, así como las ventajas y limitaciones de cada uno de ellos según lo propuesto en la **figura 1**.

El valor preventivo o terapéutico general de una determinada intervención médica solo puede ser juzgada si se compara con otra intervención alternativa (experiencia controlada), o con ninguna otra intervención (experiencia no controlada). Por naturaleza, un estudio controlado comporta una comparación estadísticamente válida entre los resultados obtenidos de un grupo tratado y otro grupo de control o testigo. Se en-

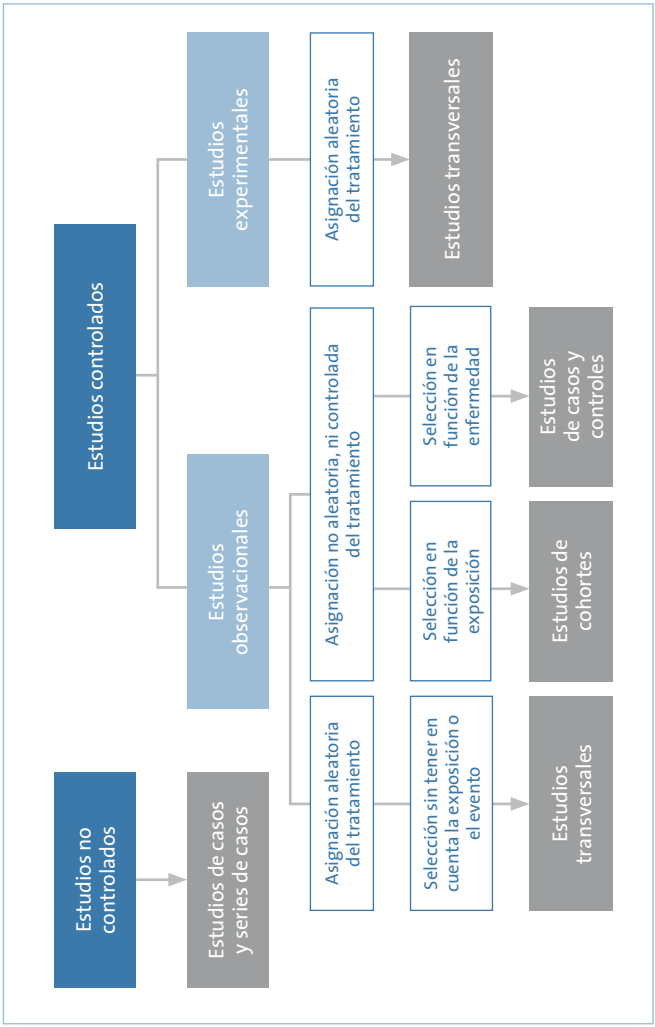

Figura 1. Tipos de diseños epidemiológicos analíticos controlados y no controlados.

tiende por *grupo de control* aquel conjunto de pacientes que recibe un tratamiento estándar, bien sea: placebo, fármaco ya conocido o simplemente, ninguna intervención, pero que en ningún caso recibe el fármaco en investigación.

Del estudio no controlado se dice que describe el curso de la enfermedad en un solo grupo de pacientes que han sido expuestos a la intervención de interés. Los estudios no controlados pueden utilizarse para generar experiencia en la mejor utilización de un medicamento, así como para estudiar efectos secundarios, cambios bioquímicos en terapias a largo plazo, tolerancia, interacciones o eficacia de los fármacos. Estos estudios que comparan el estado clínico de un grupo de pacientes antes y después de la aplicación de una intervención no nos permiten sacar relaciones de causalidad, ya que al no tener un grupo de referencia pueden conducirnos a impresiones erróneas sobre el medicamento elegido.

Por ello, y tras definir las características particulares de cada estudio, veremos en profundidad los criterios específicos de validez interna, importancia –relevancia de los resultados– y validez externa –capacidad de extrapolación de los resultados al resto de la población– para la lectura crítica de artículos de los distintos tipos de estudios controlados **(figura 2)**:

Figura 2. Tipos de estudios controlados.

1.1. Tipos de estudios en investigación clínica

1.1.1. Estudios transversales

El transversal o de prevalencia es un estudio observacional, de transición entre un estudio descriptivo y otro analítico. En él, también conocido como de encuesta, la población en estudio puede ser seleccionada de manera aleatoria sin considerar la exposición o el evento criterios de selección[1]. Se distingue por que se indaga sobre la presencia de la exposición a uno o varios factores de riesgo y la enfermedad de interés una vez conformada la población en estudio, y porque solo se hace una medición en el tiempo en cada sujeto de estudio **(figura 3)**. El número de episodios así como la proporción de sujetos con la exposición están determinados por la frecuencia con que se producen estos en la población elegible y, por tanto, quedan fuera del control del investigador[2].

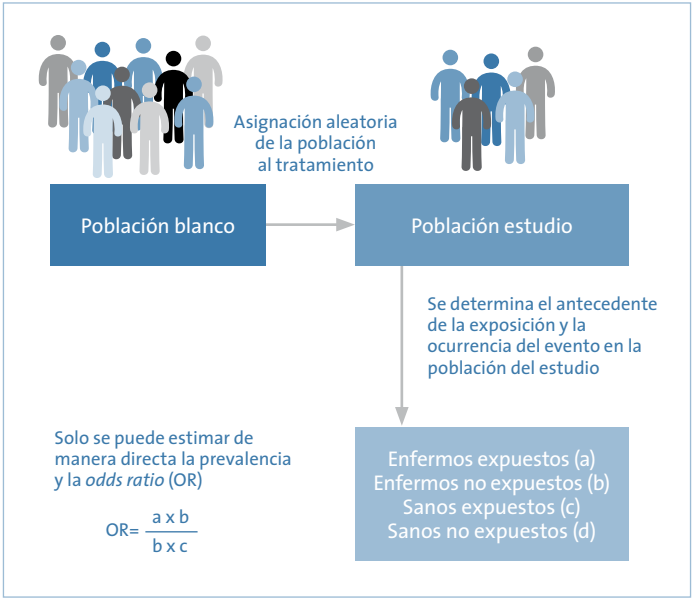

Figura 3. Esquema del diseño de un estudio transversal.

Las mediciones se realizan en un periodo único, breve y bien limitado (como una fotografía), por lo que el costo no es tan alto y al iniciar la evidencia de una asociación causa-efecto se consideran útiles en los programas de salud[3].

En la **tabla 1** se resumen las ventajas e inconvenientes de los estudios transversales.

Tabla 1. Ventajas e inconvenientes de los estudios transversales

Ventajas	Limitaciones
Fácil ejecución Relativamente poco costosos Permiten estudiar varias enfermedades y/o factores de riesgo a la vez Caracterizan la distribución de la enfermedad respecto a diferentes variables Precisan poco tiempo para su ejecución Útiles para la planificación y administración sanitaria por permitir estudiar la prevalencia de enfermedades en la población	La información sobre factores de riesgo y episodios es simultánea, lo que impide interpretar la relación causa-efecto Las medidas del impacto potencial solo permiten determinar prioridades, no metas Una población en estudio no representativa puede generar inferencias erróneas sobre una población mayor No son útiles para enfermedades raras ni de corta duración (infecciones del tracto respiratorio superior), porque las exposiciones y los resultados de la enfermedad son medidos al mismo tiempo, y las asociaciones de causa-efecto parecerán débiles, aunque sean fuertes No se pueden investigar factores de riesgo previos al estudio Posibilidad de sesgos de selección, información y confusión

1.1.2. Estudios de casos y controles

Durante los últimos años se ha desarrollado una estrategia de selección sobre la base del evento en estudio. Este tipo de muestreo se ha denominado en la literatura epidemiológica estudios de *casos y controles o casos y referentes*[4].

Son estudios observacionales (observamos, no interferimos), analíticos (buscamos una relación causa-efecto) y longitudinales (realizamos un seguimiento de los sujetos incluidos). A diferencia de los estudios de cohortes, su secuencia es de la causa al efecto. En estos estudios partimos de un grupo de individuos que presentan el efecto (enfermedad de interés) que denominamos *casos* y otro grupo en el que está ausente el efecto de interés (puede tener otros efectos o ser sano) y que denominamos *controles*. Ambos grupos se estudian de forma retrospectiva, ya que la investigación se lleva a cabo con sujetos ya enfermos con respecto a una exposición previa, y a través del interrogatorio se analizan los niveles de exposición en cada uno de ellos. Es decir, se trata de analizar el grado de asociación *(odds ratio*, cociente de las probabilidades de exposición de casos y controles) entre un factor de riesgo y un determinado evento o suceso. Precisamente por su carácter retrospectivo, los estudios de casos y controles no tienen una relación causal perfecta, ya que el evento se evalúa antes que la causa y no siempre se puede inferir que la causa antecedió al evento.

Imaginemos que queremos saber la efectividad de la vacunación antigripal en nuestra área de salud. Para ello, de los pacientes que han sufrido la enfermedad (gripe) y otros que no la han sufrido compararemos su exposición previa a la vacuna antigripal. Posteriormente, compararemos las frecuencias de exposición entre ambos grupos y, de esa manera, tendremos una aproximación de la efectividad de esta vacuna

Actualmente, los estudios de casos y controles se consideran la mejor opción para estudiar enfermedades raras o de baja prevalencia y son los más rápidos y baratos, aunque los más vulnerables a la presentación de sesgos[5].

En la **figura 4** se muestra el diseño de un estudio de casos y controles. El número de controles se usa para manejar una muestra grande y ha de ser mayor aún si se pretende investigar una enfermedad rara (prevalencia < 10%) para evadir el efecto del azar. Lo ideal es una relación 1:1 (casos:controles) y nunca se debe exceder la razón 1:4 (casos:controles).

Las principales ventajas y desventajas de este tipo de estudios se describen en la **tabla 2**.

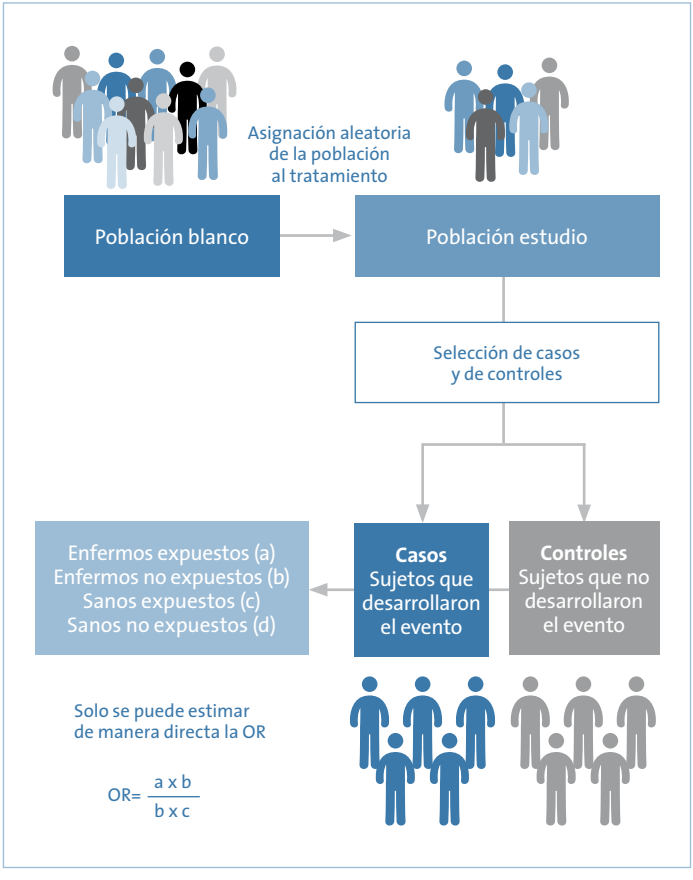

Figura 4. Diseño de un estudio de casos y controles.

Tabla 2. Principales ventajas e inconvenientes de los estudios de casos y controles

Ventajas	Limitaciones
Son estudios eficientes. Apropiados para investigar enfermedades raras o de larga latencia Más baratos que otros ensayos Permiten estudiar fenómenos que por cuestiones éticas o de rentabilidad no pueden ser analizados mediante estudios experimentales Pueden ser utilizados como estudios exploratorios de relaciones de causalidad, debido a su eficiencia (corta duración y pequeño tamaño muestral que conllevan una más rápida ejecución que otros estudios analíticos) Permiten el estudio simultáneo de varios factores de riesgo Al ser observacionales, no representan un riesgo para los sujetos del estudio	No permiten calcular la incidencia o prevalencia de la enfermedad No permiten estudiar más de un suceso o evento simultáneamente No son útiles si los criterios de diagnóstico de la enfermedad que queremos estudiar no están adecuadamente definidos Por su diseño es fácil cometer sesgos, tanto en la selección de los participantes como en la recogida de información sobre la exposición, dado que esta recogida es posterior a la aparición de la enfermedad (suceso o evento de interés)

1.1.3. Estudios de cohortes

También son estudios observacionales, analíticos y longitudinales. En estos estudios partimos de dos grupos de individuos con características en común o *cohortes*, que se clasifican a partir de su grado de exposición al objeto de nuestro estudio (factor de riesgo). Por ejemplo, unos fueron vacunados y otros no, unos fuman y otros no, y se siguen durante un determinado tiempo hasta que se produce el desarrollo de la enfermedad o evento de interés (desarrollo de una infección viral, enfermedad pulmonar...). Si al finalizar el periodo de observación la incidencia de enfermedad es mayor en el grupo de expuestos que en el de no expuestos, podremos concluir que existe una asociación entre el factor de riesgo y el evento resultante[5].

Imaginemos que queremos saber la efectividad de la vacunación antigripal en nuestra población. Para ello, elegiremos como grupo expuesto a los pacientes que han sido vacunados y como grupo no expuesto el no vacunado, y compararemos su exposición hasta que finalice el periodo habitual de aparición de gripe (diciembre/marzo). Una vez finalizado el periodo de seguimiento observamos la incidencia de gripe en cada uno de los grupos y realizamos la comparación de incidencias

En términos generales, los estudios de cohortes son los mejores diseños para describir el curso de una enfermedad o evento de salud, su pronóstico (evolución de la enfermedad asociada a un grupo de factores que se da en un grupo de pacientes frente a aquellos también enfermos, pero en los que no se dan esos factores) y la tendencia de los factores de riesgo. El diseño de cohorte es igualmente eficiente para estudiar exposiciones raras o poco frecuentes, como las exposiciones ocupacionales que se presentan en poblaciones muy reducidas de trabajadores.

La direccionalidad va de la causa al efecto y en cuanto a temporalidad, los estudios de cohorte pueden ser prospectivos y retrospectivos. Como ya ha sido definido (v. libro 3 de esta colección, apartado 2.2.4.), si son prospectivos, los sujetos del estudio se seleccionan antes de que aparezca el episodio de interés y el seguimiento se hará hasta que aparezca el suceso; mientras que si son retrospectivos, cuando se inicia el estudio tanto la exposición como el episodio han ocurrido ya. Seleccionamos los sujetos por una exposición ocurrida en el pasado (expuestos/no expuestos) y realizamos su seguimiento desde el pasado hasta la fecha actual o un pasado reciente en el que se manifestó el evento.

Entre los estudios observacionales, el de cohorte prospectivo es el más cercano al diseño experimental si está bien diseñado **(figura 5)**. Tiene un alto valor en la escala de causalidad, ya que se puede verificar adecuadamente en el tiempo la rela-

ción causa-efecto. Permite contrastar hipótesis sobre factores causales –el uso de tal vacuna impide la aparición de tal enfermedad–. La asignación de la exposición no es controlada por el investigador ni asignada de manera aleatoria, por lo que no se pueden controlar completamente las posibles diferencias entre los grupos expuesto y no expuesto en relación con otros factores asociados con la ocurrencia del episodio[5].

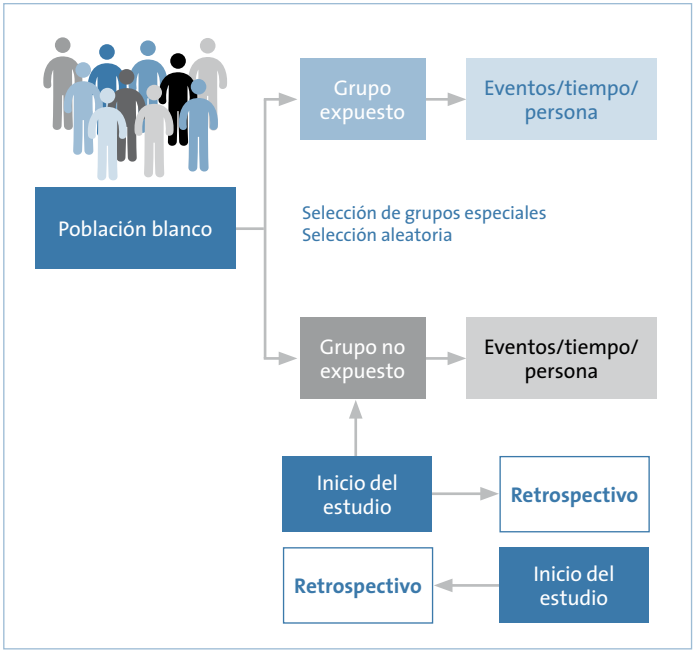

Figura 5. Diseño de un estudio de cohorte prospectivo y retrospectivo.

En la **tabla 3** se representan las principales características de los estudios de cohorte:

Tabla 3. Principales características de los estudios de cohorte

Ventajas	Limitaciones
La estimación del efecto de la exposición no está sesgada	Coste elevado
Permiten estudiar fenómenos que no pueden analizarse en estudios experimentales por cuestiones éticas o de rentabilidad	Difícil ejecución
Miden la incidencia del evento en los sujetos expuestos y no expuestos al factor de estudio	No son útiles en enfermedades raras o con periodos de latencia largos
Permiten calcular medidas de asociación (riesgo relativo, etc.) que facilitan la comprensión de resultados y de evidencias aportadas por el artículo	Normalmente, requieren un tamaño de muestra de pacientes elevado
Secuencia temporal adecuada (exposición previa al desenlace)	Los métodos y criterios diagnósticos pueden sufrir cambios con el paso del tiempo (cambios en el estado de exposición a través del tiempo)
Los prospectivos minimizan el error en la medición de la exposición, y los retrospectivos permiten acortar el tiempo de desarrollo del estudio	Posibles pérdidas durante el seguimiento
Se pueden estudiar varias enfermedades por cada exposición	Los retrospectivos facilitan la aparición de sesgos debidos a la dificultad de recogida y calidad de la información
	Pueden necesitarse grupos externos de comparación

1.1.4. Ensayos clínicos aleatorizados

Los ensayos clínicos aleatorizados son estudios experimentales que, cuando se llevan a cabo de manera adecuada, proporcionan el máximo grado de evidencia para confirmar la relación causa-efecto entre la exposición y el episodio en estudio. Se distinguen de los observacionales porque el investigador tiene control sobre la asignación de la exposición y porque esta se lleva a cabo mediante un proceso aleatorio. En el diseño de cualquier ensayo clínico se han de cumplir e incluir las siguientes características **(tabla 4)**:

Por tratarse de estudios prospectivos y longitudinales, en los que la unidad de análisis es el individuo, es muy necesario prevenir la introducción de sesgos y así incrementar la validez de los resultados[4]. La ocurrencia de sesgos en el ensayo clínico aleatorizado se minimiza mediante tres procedimientos que tienen como objetivo garantizar la comparabilidad de las intervenciones –de la exposición–, de los grupos de estudio y de los procedimientos de recogida de datos de la población estudiada. La comparabilidad de intervenciones es indicativa de la *pureza* del contraste entre los grupos experimentales y se refiere a que la única diferencia observada es atribuible al fármaco o vacuna que se está probando. Se elimina de la comparación el efecto atribuido al placebo. La comparabilidad de poblaciones se logra cuando los grupos experimentales que reciben las diferentes intervenciones son similares en todas las características que pudieran tener relación con el

Tabla 4. Características que incluir y cumplir en un ensayo clínico

Incluir	Cumplir
La ética y justificación del ensayo	Ser prospectivo
La población susceptible de ser estudiada	Utilizar un grupo de control
La selección de los pacientes con su consentimiento a participar	Llevarse a cabo en una muestra cuyo tamaño sea suficiente
Aleatorización de la intervención	Idealmente ser aleatorizado y doble ciego
Una descripción detallada de la intervención	Emplear intervenciones no sesgadas que aporten originalidad, relevancia y utilidad
Un seguimiento exhaustivo que incluya las pérdidas y los no cumplidores	Monitorización completa de los resultados que garantice la comparabilidad de la información
La medición de la variable final y la comparación de los resultados del grupo de estudio y de control	Análisis estadístico completo y con selección de la prueba estadística pertinente

episodio en estudio o con la manera en que actúa la exposición. Es decir, se trata de eliminar las variables de confusión o modificadores del efecto. Por último, la comparabilidad de la información se logra cuando se utilizan exactamente los mismos métodos de seguimiento y de medición en todos los participantes en el estudio. Esto se logra cegando o enmas-

carando a los evaluadores y a los participantes en el estudio respecto a la condición de exposición. De ahí la importancia de comprobar que el ensayo del que estamos leyendo la información es doble ciego.

Los pasos en la consecución de un ensayo clínico **(figura 6)** incluyen la definición de la población blanco, que es aquella a la que se pretende extrapolar los resultados del estudio. Mediante la aplicación de los criterios de inclusión, definimos lo que será la muestra del estudio –reclutamiento de voluntarios sanos en estudios en fase I **(figura 7)**, o de una muestra representativa de la población blanco–. Los sujetos se asignan aleatoriamente a cada rama del estudio –control o experimental–, y son seguidos en el tiempo con el fin de documentar la ocurrencia del episodio en estudio y los posibles cambios observados en otras variables de interés.

En la práctica médica actual un ensayo clínico aleatorio controlado, por ejemplo de una terapéutica contra otra, o de una vacuna frente a un tratamiento estándar, es la norma aceptada por la cual se juzga la utilidad de un tratamiento

Los ensayos clínicos tienen un papel esencial en el desarrollo de nuevos fármacos, desde que se inician los primeros estudios en animales hasta su comercialización. Una vez sintetizada la molécula, realizados los test biológicos y los primeros

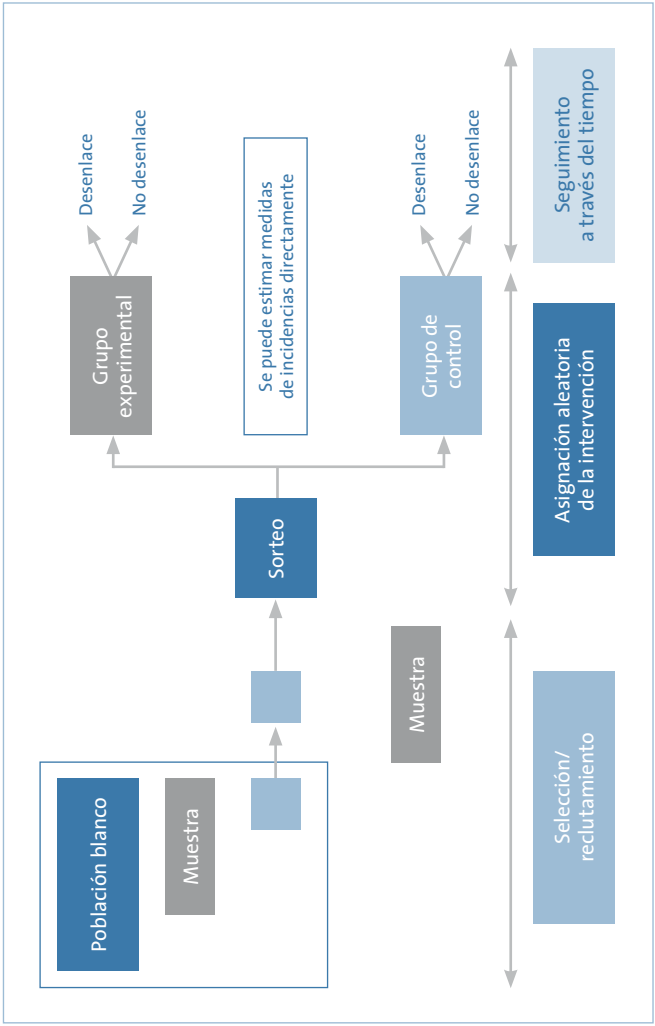

Figura 6. Diseño de un ensayo clínico aleatorizado.

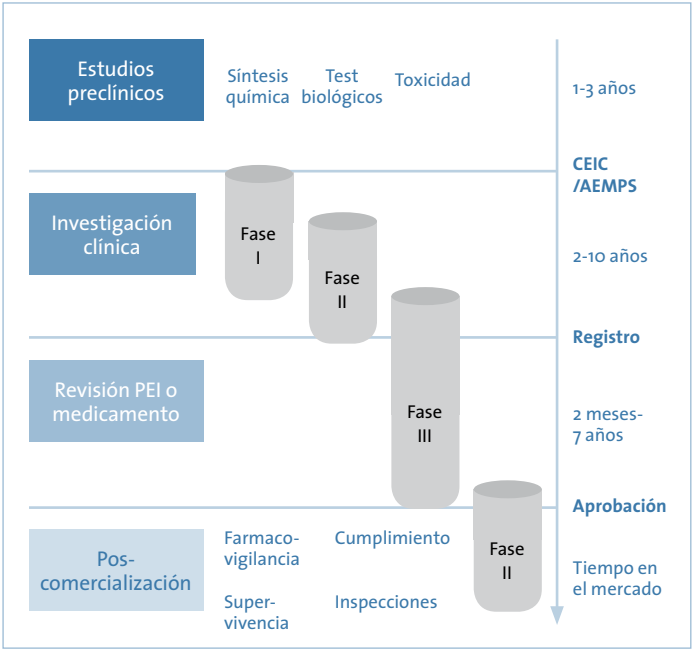

Figura 7. Fases en el desarrollo de un nuevo medicamento.

estudios de toxicidad en animales, comienzan las distintas fases de los ensayos clínicos (v. figura 7). La duración del proceso completo es variable, pero podríamos considerar 10 años como el tiempo medio que se tarda en comercializar una nueva molécula. Por el interés que representa este libro, identificaremos las diferencias en las fases de los ensayos clínicos de un medicamento y de una vacuna[6] **(tabla 5)**.

Tabla 5. Características principales de las cuatro fases de un ensayo clínico de un medicamento convencional y de una vacuna

Estudios fase I	Medicamento	Vacuna
El primer paso en la investigación de una sustancia o medicamento nuevo en el hombre. Estudios de farmacocinética y farmacodinamia. Primeros test para la evaluación preliminar de tolerancia, evidencia de la acción farmacológica, rangos y regímenes seguros de dosificación, absorción, distribución, metabolismo y excreción	No es una fase de eficacia (emplea unos 100 voluntarios sanos) Da información preliminar sobre efecto y seguridad del producto y la pauta de administración más apropiada Utiliza la misma vía de administración que en el uso clínico	El objetivo es conocer en áreas no endémicas la seguridad, tolerancia e inmunogenicidad en voluntarios humanos no inmunes. Para ello debe administrarse la vacuna a unos 20 o 30 voluntarios que den su consentimiento después de haber sido bien informados
Estudios fase II	**Medicamento**	**Vacuna**
Comprende la investigación clínica inicial del tratamiento que reúne las primeras evidencias de eficacia y seguridad	Se realiza en un número limitado de pacientes (unos 200) para estudiar la actividad biológica, el control y la profilaxis de una enfermedad Se determina la dosis adecuada y sirven para elegir fármacos con potencial para la fase III	Se aplica en vacunas cuyos resultados en la fase I han mostrado su seguridad y tolerancia El objetivo es continuar vigilando la seguridad y la tolerancia en un número mayor (100-200 voluntarios no inmunes) de individuos y valorar de manera preliminar la eficacia protectora

Estudios fase III	Medicamento	Vacuna
Incluyen un amplio rango de ensayos en los que participan numerosos investigadores para valorar la eficacia y seguridad de un fármaco nuevo bajo condiciones similares a aquellas que se puedan esperar de dicho fármaco cuando se encuentre en el mercado y considerando las alternativas terapéuticas disponibles en la medicación estudiada	Tamaño de la muestra más elevado (de varios cientos a miles) y representativo de la población general de destino del medicamento Sigue teniendo la seguridad como objetivo principal El fármaco en investigación se comparará con los tratamientos estándares, bien sea placebo o tratamientos reconocidos, a fin de confirmar evidencia de eficacia relativa	Se aplica a aquellas vacunas cuyos resultados en la fase II hayan confirmado su seguridad y tolerabilidad. La evaluación de la eficacia protectora de la vacuna en un área endémica; para ello deben realizarse ensayos clínicos controlados
Estudios fase IV	**Medicamento**	**Vacuna**
Estudios *posmarketing* con el fármaco ya autorizado, registrado y comercializado	La farmacovigilancia del producto, incluyendo la detección de efectos secundarios a largo plazo, así como posibles efectos del fármaco sobre la patología en sí misma o estudios de morbilidad y mortalidad	Monitorización de la incidencia de la enfermedad objeto de la vacunación y de los efectos secundarios de la vacuna administrada en condiciones reales a la población blanco Se establecen sistemas de vigilancia epidemiológica de la enfermedad y reacciones adversas de la vacuna

A diferencia del resto de medicamentos, los ensayos de vacunas de fase III no utilizan "una nueva medida diagnóstica y terapéutica" que ofrezca "la esperanza de salvar una vida, restablecer la salud o aliviar el sufrimiento" (investigación clínica). Sin embargo, el objetivo de la administración de una vacuna es que resulte beneficiosa para el individuo, más que "la aplicación puramente científica de una investigación médica realizada en un ser humano" (investigación biomédica no clínica). Por consiguiente, los ensayos de vacunas de fase III no corresponden a ninguna de las dos categorías definidas en la Declaración de Helsinki

Respecto a los estudios observacionales de cohortes, la principal ventaja que ofrecen los ensayos clínicos controlados es que asignan el tratamiento en prueba independientemente de determinantes pronósticos o de la selección de la muestra, de tal manera que disminuye el riesgo de sesgos de susceptibilidad diferencial (confusores) o de un muestreo. Esta y otras ventajas han sido esquematizadas a continuación **(tabla 6)**:

Tabla 6. Ventajas y limitaciones de los estudios clínicos controlados	
Ventajas	**Limitaciones**
Asignación aleatoria Disminuye el riesgo de sesgos Facilita el enmascaramiento de los individuos participantes y de sus evaluadores (doble ciego) Repetibles y comparables con otras experiencias	Coste elevado No son útiles en enfermedades raras o con periodos de latencia largos Limitaciones de tipo ético y responsabilidad en la manipulación de la exposición Dificultades en la generalización debido a la selección y/o a la propia rigidez de la intervención Puede haber sobrecontrol de las variables, que hace que los resultados se alejen de la praxis clínica diaria

1.1.5. Revisiones sistemáticas y metaanálisis

Los artículos de revisión pretenden ofrecer una actualización de los conocimientos sobre un tema clínico concreto mediante la revisión de la literatura disponible. Hay dos tipos de revisiones: narrativas y sistemáticas, aunque por la subjetividad implícita de aquellas (en la selección de artículos y análisis de su validez), solo centraremos nuestro interés en las revisiones sistemáticas.

Una *revisión sistemática* es una revisión de una pregunta formulada con claridad, que utiliza métodos sistemáticos y explícitos para identificar, seleccionar y valorar críticamente investigaciones relevantes a dicha pregunta, así como recoger y analizar datos de los estudios incluidos en la revisión[7]. Las revisiones sistemáticas son estudios pormenorizados selectivos y críticos que tratan de analizar, integrar y sintetizar la información primordial de los estudios primarios de investigación sobre un problema de salud específico. Por la forma sistemática y estructurada en la selección de artículos, su valoración crítica y la síntesis de los resultados de dichos estudios se diferencian metodológicamente de las revisiones clásicas o narrativas de la literatura científica. Por tanto, la revisión sistemática constituye un proceso de investigación más amplio, que va desde la formulación del objetivo hasta la interpretación de los resultados, incluyendo la fase de análisis estadístico cuando haya sido posible la agregación cuantitativa de los datos. Sin embargo, este tipo de publicaciones tampoco está exento de un sesgo motivado por las preferencias o resultados de los estudios realizados por el propio revisor.

Por todo ello, existe la necesidad de esfuerzo de investigación en los niveles más superiores: el *metaanálisis*. El término *metaanálisis* fue introducido por Glass en 1976 a través de la literatura en sociología y psicología, y se empezó a aplicar en medicina[8] ya a principios de 1980. Si en una revisión sistemática somos capaces de cuantificar, mediante algún índice

estadístico del tamaño del efecto, los resultados de cada estudio empírico integrado y de aplicar técnicas de análisis estadístico para extraer la esencia de dichos estudios, entonces una revisión sistemática se convierte en un *metaanálisis*. Un metaanálisis es, pues, una revisión sistemática en la que se utilizan métodos estadísticos para analizar los resultados de los estudios integrados en ella[9].

Los metaanálisis y las revisiones sistemáticas han surgido de la necesidad de sintetizar, valorar y poner al día la información médica, tratando de encontrar la mejor evidencia científica ante una cuestión concreta y, al mismo tiempo, presentando esta evidencia de una forma clara, sintética y de fácil comprensión. Ahora bien, por su mayor nivel de cuantificación y rigor, el metaanálisis es el tipo de revisión que nos ofrece las evidencias más válidas de un problema médico

Al aplicar técnicas estadísticas, un metaanálisis nos permite responder a preguntas tales como: ¿cuál es la magnitud del efecto global de los diferentes tratamientos?, ¿son homogéneos los resultados de eficacia alcanzados por los diferentes tratamientos? En caso de que no sean homogéneos, ¿qué

factores son los que pueden explicar esa heterogeneidad en los resultados? y ¿es posible formular un modelo explicativo sobre dicha heterogeneidad en los resultados? Para dar respuesta a estas preguntas, un metaanálisis implica desarrollar los pasos típicos de una revisión sistemática y aplicar técnicas estadísticas de integración: *a)* formulación del problema; *b)* identificación y selección de la literatura más relevante; *c)* codificación de los estudios; *d)* análisis estadístico e interpretación, y *e)* publicación del metaanálisis. La descripción de cada etapa se presenta en la **tabla 7**:

Tabla 7. Etapas de una revisión sistemática

(1) Formulación del problema	El corazón de toda revisión sistemática de alta calidad es la formulación clave de la pregunta primaria Una mala pregunta es la antesala de una mala respuesta o de la ausencia de respuesta
(2) Identificación y selección de la literatura más relevante	Una buena estrategia de búsqueda ayuda a minimizar los sesgos, por reducir la no recuperación de estudios relevantes Una revisión sistemática de calidad implica la necesidad de una buena búsqueda manual La base de datos bibliográfica constituida por todos los estudios será la piedra angular de la revisión

Tabla 7. Etapas de una revisión sistemática *(continuación)*	
(3) Codificación de los estudios: registro de las variables moderadoras del estudio	Variables del tratamiento aplicado Variables de los participantes del estudio (edad, sexo, gravedad de la enfermedad, etc.) Variables de contexto o lugar donde se llevó a cabo la intervención (unicéntrico, multicéntrico, etc.) Variables metodológicas o de diseño del estudio (tipo de ensayo clínico, tamaño de la muestra, mortalidad experimental, tipo de análisis estadístico, etc.) Variables extrínsecas o ajenas al proceso científico de la investigación, pero que pueden influir directamente en la presentación de los resultados (fuente de la publicación, especialidad de los autores, año de publicación)
(4) Análisis estadístico e interpretación de los resultados	Cálculo de un índice cuantitativo que pone en la misma métrica todos los estudios Se suele usar la diferencia entre las medias (μ_1-μ_2) tipificadas respecto a la desviación típica conjunta También se puede usar la diferencia entre proporciones, RR y OR
(5) Publicación del metaanálisis	Han de cumplir las normas de cualquier otro estudio empírico (introducción, métodos, resultados, discusión y conclusiones)

El metaanálisis proporciona una metodología científica, rigurosa, transparente, sistemática y cuantitativa para el desarrollo de las revisiones de la investigación, necesaria para la acumulación eficiente del conocimiento y la optimización de las decisiones profesionales en alergología y otras áreas de la salud[10]. De forma concreta y detallada, podríamos destacar las siguientes ventajas y limitaciones **(tabla 8)**:

Tabla 8. Ventajas y limitaciones de un metaanálisis

Ventajas	Limitaciones
Permite tratar grandes cantidades de información Maneja tamaños de muestra muy elevados, resultado de acumular las muestras de todos los estudios Cumple con todas las características del método científico y destaca su reproducibilidad Conclusiones precisas, objetivas y contrastables (análisis estadístico de los resultados cuantitativos) La estadística del metaanálisis permite el análisis de heterogeneidad en los resultados de los estudios	Sesgo de publicación o tendencia a publicar los estudios estadísticamente significativos Las limitaciones y deficiencias metodológicas de los estudios que se integran en el metaanálisis restan validez a sus conclusiones La heterogeneidad excesiva en la metodología, diseño, operatividad de las variables y conceptualización de los estudios pueden dificultar enormemente su combinación

1.2. Lectura crítica de artículos con diseño transversal

Los estudios transversales permiten medir la prevalencia de una enfermedad o su condición de interés. Se definen como estudios epidemiológicos transversales aquellos en que la medida de la exposición y de su efecto se llevan a cabo en el mismo momento. Habitualmente, tienen la finalidad de estimar la frecuencia (prevalencia) de un estado o situación, o una medida de asociación *(odds ratio*, o razón de prevalencias) entre una exposición y un resultado, aunque no se disponga de seguimientos temporales[11]. Lo más complicado de un estudio transversal es dilucidar correctamente la naturaleza de las asociaciones que se observan: la cuestión clave es si la exposición precede o es posterior al efecto.

A continuación, se enumeran los criterios principales para la lectura crítica de este tipo de artículos:

1. ***¿Cuál es la población de referencia del estudio?*** Las primeras preguntas que nos hacemos siempre van dirigidas a definir con claridad la población base o diana del estudio.

2. Respecto a la **muestra** estudiada: ***¿es representativa la población de referencia? ¿Es necesario que lo sea?*** Nos interesa saber si la muestra estudiada es representativa de la población general o de una determinada subpoblación (p. ej., de los enfermos con dermatitis atópica o de los tra-

bajadores del sector hospitalario), dado que se pretende hacer la inferencia sobre esa subpoblación a partir del estudio de un número lo más reducido posible de sujetos. En ocasiones, no es necesario que la muestra elegida sea representativa de ninguna población, como en los estudios de series de casos o en estudios descriptivos asistenciales[2].

¿Están claramente definidos los métodos de muestreo y el tamaño de la muestra? Cuando se precisa de una muestra representativa, se debe valorar el tamaño de la misma y el método de muestreo utilizado. *¿Cuál es el grado en que difieren la muestra originalmente seleccionada y la estudiada finalmente?* Este es un criterio muy importante para evaluar la validez interna del estudio. Debemos conocer si la seleccionada al principio es muy distinta de la que realmente se estudia, así como la proporción de no respuesta y si el motivo pudiera relacionarse con el tema de estudio, que puede estar condicionando la participación. Por eso nos encontramos artículos en los que, además de incluir la proporción o tasa de respuesta, se incluyen datos que comparan algunas características basales o de referencia (sexo, edad, nivel de estudios o cualquier otra variable sociodemográfica) de los sujetos que participan y de los que no llegan a participar del estudio.

3. *¿Cuáles han sido los métodos para medir la exposición (variable independiente) y la condición de interés (variable dependiente)?* Se han de definir y medir claramente las exposiciones –duración, dosis, instrumentos y cuestionarios validados–, así como la condición de interés, que puede ser una enfermedad –criterios diagnósticos, medidas, etc.–, un estilo de vida –ser fumador, sedentarismo– o una condición social o personal. Es importante fijarse en la estandarización de los métodos de medida y en su correcta aplicación en todos los sujetos estudiados, sin sesgos e independientemente entre exposiciones y enfermedades, con el mismo esfuerzo en toda la muestra. También se han de incluir preguntas sobre la temporalidad de las exposiciones y de las enfermedades.

4. **En cuanto a la valoración de los resultados**, en los estudios transversales se presentan las prevalencias (proporciones). Como cualquier medida de frecuencia, y también de asociación, hay que exponer la estimación puntual y el intervalo de confianza, generalmente del 95%. Los resultados han de estar ajustados por las posibles variables de confusión. En muchos casos, es pertinente explicar el tipo de ajuste o estandarización para controlar el efecto de una tercera variable, por ejemplo la edad. Finalmente, habrá que revisar la interpretación de la importancia de los resultados –magnitud y precisión–, incluyendo su significado no causal, es decir, solo como medida de asociación.

5. **En la sección de interpretación y discusión** de los resultados debemos identificar claramente el razonamiento de los autores sobre las ventajas y limitaciones del estudio –sesgos de selección, de información y de confusión–, y cómo estas limitaciones contribuyen a la explicación de los resultados. Es muy importante que, en el caso de que se establezcan asociaciones, se discutan a partir de las hipótesis planteadas y de la temporalidad de las exposiciones y de sus efectos.

1.3. Lectura crítica de artículos de casos y controles

Aunque los estudios de casos y controles son inferiores en cuanto a calidad respecto a los estudios de cohortes y son susceptibles a la acción de sesgos sistemáticos, un estudio de este tipo bien diseñado presenta muchas ventajas y muy pocos inconvenientes en comparación con un estudio de cohortes que, por otra parte, resulta muy costoso. Por su menor coste en tiempo y recursos, el estudio de casos y controles es el que nos encontraremos con más frecuencia en la literatura médica a la hora de investigar los factores de riesgo asociados a una determinada enfermedad[12].

Hay una serie de aspectos que debemos revisar en cuanto a la selección de los casos y de los controles, para considerar que la evidencia aportada es consistente **(tabla 9)**:

Tabla 9. Aspectos que revisar para la selección de casos y controles	
Respecto a los casos	Identificar la fuente y base del estudio de los que provienen los casos Casos incidentes/prevalentes: conviene que prevalezcan los incidentes para evitar el sesgo de supervivencia Descripción de los criterios diagnósticos de la enfermedad Definición clara de los criterios de inclusión y exclusión Descripción de las pérdidas: casos identificados frente a casos incluidos finalmente
Respecto a los controles	Identificar la fuente y base del estudio de los que provienen los controles ¿Tienen la misma probabilidad de exposición que los casos? Comprobar que el proceso asistencial es el mismo que el de los casos, en el supuesto de que enfermaran ¿Hay sesgos de selección? ¿Son estos idénticos a los sesgos de selección de los casos? En caso de que haya más de un grupo de control, ¿qué diferencias hay entre ellos y qué grado de influencia pueden tener sobre los valores de OR? Definición clara de los criterios de inclusión y exclusión

Un aspecto especialmente relevante de la **validez** de los estudios de casos y controles es **la selección de los grupos que comparar**. Casos y controles han de representar la misma población fuente o de origen. La selección del grupo de control ha de tener la misma procedencia (p. ej., el mismo hospital o centro de salud) que los casos y representar de un modo adecuado la población de la que proceden los casos. Dicho de otro modo, casos y controles han de ser similares en todo, excepto en la ocurrencia del episodio que estamos estudiando[5]. El objetivo de este tipo de estudios consiste en que los controles deben permitirnos estimar la distribución de la exposición o exposiciones de interés en la población de referencia de la que proceden los casos, de manera que si un control enfermara podría convertirse en un caso potencialmente elegible.

Como lectores de un artículo de casos y controles, pondremos nuestra atención en el grado de adecuación de la población de casos, lo que nos permitirá valorar críticamente la identificación correcta o no de los controles. A continuación, los propios criterios de selección de los casos (recientes/prevalentes, inclusión/exclusión o de diagnóstico) nos permitirán valorar la validez del estudio y sus conclusiones

Por su solidez, interesa que en el estudio predominen los casos incidentes (recientes) sobre los prevalentes. Cuando se trabaja con casos prevalentes las conclusiones a las que podemos llegar son más inciertas, por haber un mayor sesgo de supervivencia. Por ejemplo, un estudio que incluya una cardiopatía congénita está, por definición, basado en casos prevalentes. Este diseño no nos permite sacar la información válida sobre los factores que determinan la incidencia real de este episodio.

Los criterios diagnósticos para la asignación de casos y controles deben ser una garantía de calidad, de modo que un caso no va a ser un falso positivo de la enfermedad, pero tampoco un control es un caso falso negativo.

Como en cualquier otro diseño de investigación, **la medida de la exposición** y otros factores de interés son muy importantes. En los estudios de casos y controles la exposición es retrospectiva, lo que hace que los errores de clasificación respecto a la variable sean más probables. Los autores han de aportar toda la información relevante sobre la medida de la exposición (criterios, marcadores, duración, dosis, instrumentos y cuestionarios) para que podamos caracterizar adecuadamente la validez de su medida y otras variables de interés. Por otra parte, la exposición se ha de medir sin sesgos respecto a la enfermedad estudiada y se debe destinar el mismo esfuerzo para reunir información sobre la exposición en casos y controles y ha de ser previa a la aparición de la enfermedad.

Los resultados deben presentarse de manera que seamos capaces de valorar su precisión. Las estimaciones de cada *odds ratio* han de venir acompañadas de sus correspondientes intervalos de confianza. Es necesario que los resultados se controlen por las variables de confusión relevantes y, en caso de no aportar significación estadística, que se justifique que el tamaño de la muestra era correcto. Asimismo, por tratarse de estudios que han de aportar evidencias sobre la relación causa-efecto entre la exposición o exposiciones de interés y la enfermedad presente en los casos, es muy importante encontrar en la discusión una valoración de las consideraciones para la causalidad propuesta por Hill en 1965 (v. libro 3 de esta colección, apartado 2.3.4.). Finalmente, los autores del estudio tendrán que discutir las ventajas y limitaciones del mismo, describiendo los potenciales sesgos de selección, información y confusión y su efecto sobre los resultados.

1.4. Lectura crítica de artículos de cohortes

Se utiliza para estudiar un grupo de individuos con características comunes (cohortes) que se siguen conjuntamente en el tiempo. Son estudios observacionales y analíticos en los que un grupo de individuos es definido sobre la base de la presencia o ausencia de una exposición o factor de riesgo hipotético para una enfermedad (p. ej., unos toman el fármaco y otros no, unos son alérgicos y otros no), para luego seguirlos duran-

te un periodo de tiempo y evaluar la incidencia del resultado (la enfermedad).

Los elementos comunes a todos los tipos de estudios de cohorte son: *a)* la identificación de la población en estudio, o cohorte, o de personas expuestas al factor de interés; *b)* la identificación de la población de control *(población de referencia)*; *c)* el seguimiento de las cohortes en el tiempo, y *d)* la comparación de las tasas de enfermedad entre la cohorte y la población de referencia.

Dicho esto, lo primero que hemos de hacer durante la lectura crítica del artículo correspondiente a un estudio de cohortes es precisamente eso: saber si se describen bien las características de la cohorte. Es decir, su origen y tipo, el tiempo de seguimiento y los criterios de inclusión y exclusión de los individuos.

Según se defina el periodo de seguimiento de los participantes de un estudio de cohortes, hablamos de cohortes **prospectivas o retrospectivas**. Los estudios de cohortes son habitualmente prospectivos: el investigador elige una muestra de individuos que no presentan el episodio o la enfermedad de interés, de modo que: *a)* mide los factores que podrían predecir la aparición del desenlace en cada uno de los individuos, y *b)* efectúa un seguimiento de este conjunto de personas mediante encuestas o exámenes periódicos para detectar

la aparición del episodio en el tiempo. Existe también la posibilidad de realizar un estudio de cohorte retrospectivo: la diferencia es que la selección del grupo de sujetos, las mediciones, el seguimiento y el desarrollo del episodio se produjeron en el pasado. Este estudio es posible si se dispone de datos adecuados sobre los factores de riesgo y los resultados o episodios en estudio. Por ejemplo, un investigador diseña un estudio en el que se propone determinar si se quiere saber en un grupo de pacientes que han sufrido infarto del miocardio agudo cuáles eran los hábitos dietéticos, la actividad física sistemática que realizaban y si fumaban antes de sufrir el ataque cardiaco.

Con relación al tipo de población, las cohortes construidas pueden ser **fijas** *(cerradas)* y **dinámicas**[13]. Cerradas o fijas son las que, por diseño, no consideran la inclusión de población en estudio más allá del periodo de reclutamiento fijado por los investigadores, por ejemplo: graduados MIR en alergología por la Universidad Complutense de Madrid, promoción de 2014. Dinámicas son aquellas cohortes que consideran la entrada y salida de nuevos sujetos de estudio durante la fase de seguimiento, por lo que el número de miembros puede variar a través del tiempo. Los participantes entran o salen de la cohorte cuando cumplen criterios de elegibilidad, incorporando la aportación años-persona desde el momento de inclusión en el estudio; frecuentemente están definidos por unidades geográficas y grupos poblacionales, como pobla-

ción con diagnóstico de asma alérgica en un hospital durante el periodo de 2000 a 2014.

Cuando la entrada a la cohorte es única, se tiene el estudio de cohortes fijas. Si hay entradas diferentes a lo largo del seguimiento, se tiene el estudio de cohortes dinámicas. En algunos casos, se incluye una cohorte única y se clasifican los individuos en función de su estado de exposición *(comparación interna)*. Este podría ser el caso del estudio del efecto del tabaco en un grupo común de sujetos (p. ej., estudiantes de medicina) en el que se los clasifica a posteriori a partir de su consumo basal de tabaco. En otros casos, se utiliza una cohorte de individuos expuestos, utilizando la población general de origen de la cohorte como grupo de referencia (*comparación con la población general)*. Podría estudiarse la mortalidad por diferentes tipos de cáncer entre los trabajadores de un determinado sector industrial y compararse con las tasas de mortalidad por cada una de estas variantes de la enfermedad en la población general. También puede ocurrir que lo que leamos compare dos cohortes definidas (por ejemplo, usuarios y no usuarios del teléfono móvil) y seleccionadas deliberadamente en función de su estado de exposición al factor de interés *(comparación externa)*.

En los estudios de cohortes, los autores del trabajo han de presentar toda la información necesaria para evaluar la validez del criterio y de los resultados obtenidos de la **medida de la exposición**. La medida de la exposición se puede basar en

marcadores biológicos o ambientales, cuestionarios u otros tipos de registros o instrumentos.

Aunque se reconoce que los estudios de cohorte representan un diseño menos sujeto a sesgos en comparación con otros estudios observacionales[14], no es menos cierto que se deben tener en consideración algunas fuentes que pueden distorsionar los resultados que se deriven de ellos. En los estudios de cohortes hay sesgos de selección e información que deben ser considerados con rigor, sobre todo en lo que se refiere a pérdidas de seguimiento, a cómo se obtiene la información sobre la exposición estudiada y al modo en que se determina en la población en estudio la ocurrencia de la enfermedad o condición de interés durante el seguimiento. Por lo que se refiere a sesgos de confusión, en los estudios de cohorte es importante considerar factores que se asocien independientemente tanto con la exposición como con la condición o episodio estudiados que no sean pasos intermedios en el proceso causal, ya que estos pueden hacer aparecer una asociación ficticia entre la exposición y el factor en estudio[15]. Los autores han de garantizar que la exposición es previa a la aparición de la enfermedad.

Por otra parte, idealmente, la experiencia de enfermedad se debe cuantificar de forma completa y válida en todos los sujetos de la cohorte[12]. Esto es crucial para garantizar la validez del estudio. Asimismo, para evitar sesgos diferenciales, es importante que los métodos aplicados en las cohortes en com-

paración y en sus individuos sean equivalentes. Es decir, se ha invertido el mismo esfuerzo en reunir la información sobre la enfermedad en todos los miembros de la cohorte.

La definición del **periodo de seguimiento** para todos los miembros de la cohorte es clave. El investigador selecciona un grupo de sujetos expuestos y otros no expuestos, y los sigue en el tiempo para comparar la incidencia del episodio (enfermedad, tasa de mortalidad). Es indispensable considerar que para poder analizar adecuadamente un estudio de cohorte se necesitará información sobre la fecha de inicio, en que ocurran los episodios y de terminación del estudio, así como la información completa de los datos de los participantes, la escala de medición y el motivo de terminación del seguimiento (pérdida, muerte u ocurrencia del episodio en estudio)[13]. El periodo de seguimiento puede abarcar años, meses, semanas o días, dependiendo de la frecuencia del episodio estudiado. Dos momentos definen el periodo de seguimiento: *el examen inicial* (medición basal) y el *final del seguimiento*. El inicio del seguimiento depende del tipo de cohorte: si es cerrada o dinámica, ya que en el caso de esta última, el inicio del seguimiento se define para cada participante a través de un largo periodo de tiempo. El seguimiento, dependiendo del episodio de interés, puede ser activo o pasivo. Activo es aquel en el que se utilizan contactos repetidos por diversos medios, nueva entrevista y obtención de muestras, cuestionarios autoaplicables o llamadas telefónicas. El pasivo es el que se realiza

mediante búsqueda sistemática de sistemas de información en registros preestablecidos (registros de cáncer, hospitalarios, registro civil, entre otros). Sea como sea, lo más importante es valorar que el seguimiento es idéntico para todos los miembros de la cohorte y suficiente para observar el suceso de interés.

Los criterios relativos a la lectura crítica de los resultados y discusión de los artículos de cohortes no difieren mucho de los de casos y controles. En los estudios de cohortes se presentan las estimaciones de frecuencia ajustadas al diseño de la cohorte en cuestión (tasas de incidencia, incidencia acumulada, tasas estandarizadas por edad) y las estimaciones de asociación (riesgos relativos, razones de incidencia, razones de mortalidad) ya sean crudas o ajustadas por las variables de confusión.

Para finalizar, en la discusión, el lector tendrá que encontrar la comparación de los resultados más relevantes con las evidencias extraídas de estudios previos. También se ha de incluir una reflexión sobre los criterios de causalidad (dosis-respuesta, plausibilidad biológica).

A diferencia de los estudios de casos y controles, la temporalidad no es un problema en los estudios de cohortes, ya que durante el seguimiento de los grupos de estudio se obtiene información directa sobre la relación temporal entre la exposición al factor de estudio y el desarrollo de la enfermedad.

1.5. Declaración STROBE

En investigación médica hay muchas cuestiones que se evalúan a través de estudios observacionales[16]. Gran parte de la investigación relacionada con la etiología de las enfermedades se lleva a cabo mediante los estudios de cohortes, estudios de casos y controles o estudios transversales. Los estudios observacionales, cuya lectura crítica acabamos de ver en los tres apartados anteriores, tienen también un papel importante en la investigación de los efectos beneficiosos y perjudiciales de las intervenciones médicas[17].

La investigación debe ser comunicada de forma transparente para que sea creíble, pero su credibilidad dependerá de la evaluación crítica que hagamos sobre las fortalezas y debilidades de diseño de cada estudio. La realización de recomendaciones sobre la comunicación de la investigación puede ayudar a mejorar la calidad de la misma. Se estableció así una red de especialistas en metodología, investigadores y editores de revistas para el desarrollo de recomendaciones para la comunicación de la investigación observacional: la declaración STROBE *(Strengthening the Reporting of Observational Studies in Epidemiology)*. Está constituida por una lista de puntos que hay que tener en cuenta en la comunicación de los tres diseños más importantes de la epidemiología analítica observacional: los estudios de cohortes, los estudios de casos y controles y los estudios transversales[18].

Se trata de una lista con 22 puntos que se consideran esenciales para la comunicación adecuada de los estudios observacionales que fue desarrollada para ayudar a los autores a redactar sus estudios observacionales analíticos, a los editores y a los revisores que consideran la posible publicación de estos artículos y también para ayudar a los lectores que evalúan de forma crítica los artículos publicados.

El listado de todos los puntos de STROBE son los que figuran en un artículo que explica y detalla la elaboración de cada punto de la lista y se ofrece el contexto metodológico y ejemplos reales de comunicación transparente[19]. La lista de puntos STROBE se debe utilizar preferiblemente junto con ese artículo (gratuito en las páginas web de las revistas PLoS Medicine [http://www.plosmedicine.org]). En la página web de STROBE (http://www.strobe-statement.org) aparecen las diferentes versiones de la lista correspondiente a los estudios de cohortes, a los estudios de casos y controles y a los estudios transversales (modificado de Vandenbroucke et al.[18]).

Estos puntos reflejan la estructura convencional de cualquier artículo: título y resumen (punto 1), introducción (puntos 2 y 3), metodología (puntos 4-12), resultados (puntos 3-17) y discusión (puntos 18-21), así como otros apartados relevantes (punto 22 relativo a la financiación). Los tres diseños tienen 18 puntos comunes, mientras que 4 puntos son específicos de los diferentes diseños (puntos 6, 12, 14 y 15), y se proporcionan versiones diferentes (ya sean totales o parciales) para

estos puntos según el diseño específico. Para algunos (marcados con un asterisco), la información se debe presentar por separado para los casos y para los controles en los estudios de casos y controles, y para los participantes expuestos y no expuestos en los estudios de cohortes y transversales. Aunque en este libro se muestra una única lista, realmente se elaboraron listas distintas para cada uno de los tres tipos de diseños (disponibles en **http://www.strobe-statement.org**).

Ahora bien, la declaración STROBE no se ha desarrollado como una herramienta para evaluar la calidad de los estudios observacionales publicados. Y tampoco debe interpretarse como un intento de recomendar la comunicación de los estudios observacionales en un formato rígido. Los puntos de la lista deberían ser abordados con detalle y claridad suficiente en alguna parte de un artículo, pero el orden y el formato de presentación de la información dependen de las preferencias de los autores, del estilo de la revista y de las tradiciones existentes en ese campo de la investigación. Si los autores y las revistas adoptaran las recomendaciones STROBE, aspectos como los factores de confusión, los sesgos y la generalización de los resultados serían más transparentes y ayudarían a calmar el deseo excesivamente entusiasta de comunicar nuevos hallazgos, tanto por parte de la comunidad científica como de los medios de comunicación social[20], además de mejorar la metodología de los estudios a largo plazo.

1.6. Lectura crítica de artículos de estudios experimentales

Los estudios experimentales comparan dos o más grupos de sujetos sometidos a intervenciones diferentes (farmacoterapia, práctica preventiva, intervención quirúrgica, etc.) y el tipo más común son los ensayos clínicos aleatorios.

El diseño de los ensayos clínicos aleatorios viene determinado por el control de la variable a estudio y la asignación aleatoria de los sujetos a cada intervención. Los grupos estudiados son teóricamente comparables en todas las variables excepto en la intervención que reciben.

Las preguntas clave de la lectura crítica de los ensayos clínicos aleatorios han de ir dirigidas a analizar la imparcialidad de las comparaciones: *a)* ¿se asignaron aleatoriamente las intervenciones?; *b)* ¿se analizaron todos los sujetos que fueron incluidos en el estudio?, y *c)* ¿se mantuvo un diseño ciego respecto a la valoración de los resultados?

Los grupos de estudio deben ser comparables en todo, excepto en la intervención que reciben, de modo que si existen diferencias entre uno y otro grupo, serán atribuibles a la intervención en sí.

Desde luego que una cuestión complicada de los estudios experimentales es dilucidar correctamente la asignación aleato-

rizada de los individuos a cualquiera de las ramas del estudio y la posterior comparabilidad entre los grupos de estudio. Por ello, es importante que la lectura de cualquier artículo que recopile la investigación clínica llevada a cabo bajo el marco de ensayo clínico se haga con un criterio de valoración crítica. De hecho, la lectura crítica de un ensayo clínico debe incluir: *a)* la valoración de su validez interna; *b)* la valoración de la importancia de los resultados, y *c)* la valoración de su aplicabilidad.

1.6.1. Sobre la validez interna del estudio. ¿Son válidos los resultados de un ensayo clínico?

Una cuestión inicial que hay que plantearse es discernir si el efecto observado es debido a la intervención y no a otras variables que pueden llevar a una falsa conclusión. Es crucial identificar este asunto al inicio de la lectura del artículo y puede determinar nuestra decisión de continuar con la lectura del trabajo[21].

Hay tres criterios fundamentales que todo ensayo clínico ha de reflejar en su resumen: que se hayan definido claramente los objetivos del estudio, que la asignación de los sujetos haya sido aleatoria y que se haya seguido a todos los pacientes incluidos al inicio y se haya llevado a cabo un análisis por intención de tratar. Otros aspectos secundarios incluyen: el diseño ciego, la similitud de las características basales de los grupos y la igualdad de tratamiento de los distintos grupos, excepto en la intervención que se pretende estudiar[22].

- *Clara definición de los objetivos. ¿Se orienta el ensayo a una pregunta claramente definida?* El ensayo clínico debe definir claramente sus objetivos y detallar explícitamente: las características de los pacientes incluidos, la intervención estudiada, la intervención o intervenciones frente a las que se compara y las variables con las que se evalúa su eficacia.

- *Asignación aleatoria. ¿Fue aleatoria la asignación de los pacientes al tratamiento? ¿Se ocultó la asignación aleatoria?* La asignación aleatoria o aleatorización (definida en el libro 3 de esta colección, apartado 2.2.4.) nos asegura una distribución equilibrada de las variables pronóstico entre los grupos de estudio, excluyendo las diferencias de resultado atribuibles a variables no controladas. La asignación aleatoria implica que todos los participantes en el ensayo tienen la misma probabilidad de entrar en cualesquiera de los brazos del estudio. Y su importancia radica en que este diseño es el que más se aproxima a la formación de grupos de pacientes similares, no solo respecto a variables conocidas como pueda ser la gravedad de la enfermedad, sino también respecto a otras variables de influencia desconocida.

- *Seguimiento adecuado. ¿Se tuvieron en cuenta todos los pacientes en el momento del análisis?* Los grupos que forma la aleatorización serán comparables siempre y cuando el seguimiento que se haga de ellos también sea similar.

La verificación del correcto análisis de todos los pacientes participantes en el estudio incluye la comprobación de un seguimiento completo y un análisis de los resultados por intención de tratar. Durante el desarrollo de los ensayos clínicos siempre se producen pérdidas de pacientes (abandonos, retiradas, pérdida de seguimiento, desviaciones del protocolo). Sin embargo, todos los pacientes incluidos en un ensayo clínico deben ser tenidos en cuenta en la conclusión del mismo. Todas las pérdidas posaleatorización deben aparecer reflejadas en el informe del ensayo clínico. De lo contrario, los resultados obtenidos no serían comparables.

- *Análisis por intención de tratar. ¿Se analizaron los pacientes en el grupo al que fueron aleatoriamente distribuidos?* En los ensayos clínicos, como en la práctica diaria, se producen con cierta frecuencia desviaciones del protocolo, cuando un paciente que fue asignado aleatoriamente a recibir un tratamiento deja de tomarlo, lo sigue parcialmente o recibe en realidad, por distintos motivos, el tratamiento de otro grupo. Ante estas circunstancias cabrían teóricamente tres posibilidades: *a)* excluir al paciente del análisis final; *b)* analizarlo en el grupo del tratamiento que realmente siguió, aunque fuera asignado a otro *(análisis por tratamiento)*, y *c)* analizarlo en el grupo de tratamiento al que fue asignado aleatoriamente, aunque realmente no lo recibiera *(análisis por intención de tratar)*. Aunque pueda resultar paradójico, el enfoque correcto es anali-

zar al paciente en el grupo en el que fue aleatoriamente asignado. Las técnicas de análisis por intención de tratar tienen por objeto evaluar los resultados de los sujetos en el grupo al que fueron asignados con independencia de que lo recibieran o completaran el estudio.

A través del análisis por intención de tratar, por un lado, se conservan las propiedades conseguidas con la asignación aleatoria (la homogeneidad de los grupos), y por otro, se aproxima a la realidad de la práctica clínica (efectividad), en la que los cambios o incumplimiento de los tratamientos son habituales.

- *Diseño ciego. ¿Se mantuvieron ciegos respecto al tratamiento, tanto los médicos como los pacientes, durante el estudio?* Es bien conocido por todos que las expectativas de los pacientes y de los investigadores pueden influir en la evaluación de los resultados. Este problema se puede prevenir con las técnicas de enmascaramiento o ciego, que tratan de evitar que tanto pacientes como investigadores *(doble ciego)*, así como evaluador y/o analistas *(triple ciego)* conozcan el tratamiento asignado. Esta técnica está bien desarrollada mediante el uso de un medicamento o procedimientos placebo, aunque no siempre es posible aplicarla. En ese caso, se ha de evaluar la posibilidad de que los pacientes y/o los observadores puedan influir en la respuesta.

- *¿Fueron tratados todos los grupos de forma similar, aparte de la intervención experimental?* Se debe comprobar que todos los pacientes incluidos en el ensayo han sido tratados de forma similar, salvo en la intervención que se va a estudiar. Si los pacientes de alguno de los grupos recibieron mayor atención, fueron visitados con mayor frecuencia o tomaron distintos tratamientos complementarios, los resultados pueden verse sesgados.

1.6.2. Valoración de la importancia del estudio. ¿Cuáles son los resultados?

Acabamos de ver que los grupos del estudio deben ser comparables tras la asignación aleatoria y durante el seguimiento. Este principio debe quedar bien definido en el artículo mediante tablas que describan la distribución de las variables basales y pronósticas de los grupos de estudio. Estas tablas deben incluir las frecuencias relativas de las variables cualitativas y ordinales, las medidas de tendencia central (media, mediana) y dispersión (desviación estándar) de las variables cuantitativas continuas y un test de significación estadística sobre la hipótesis nula de igualdad entre los grupos de estudio. No todos los ensayos clínicos utilizan como criterio de medición variables dicotómicas. En ocasiones, el efecto se refiere a una variable cuantitativa. En estos casos, la medida del efecto de la intervención podrá estimarse calculando la dife-

rencia entre los valores medios de dichas variables en los distintos grupos de estudio (diferencia de medias o medianas).

Ahora bien, cuando estemos decidiendo si un nuevo fármaco es más eficaz que la intervención estándar o placebo, debemos basarnos en las pruebas de significación estadística, el tamaño del efecto y la precisión de la estimación para considerar su relevancia clínica y/o sociosanitaria

A partir de este momento, será necesario valorar la importancia de los resultados del estudio. La importancia clínica de un tratamiento se valora mediante la magnitud del efecto hallado y la precisión de los resultados[22].

- *¿Cuál es la magnitud del efecto del tratamiento estudiado?* Conocer la magnitud del efecto pasa por saber evaluar las diferentes estrategias para cuantificar la fuerza de una asociación entre un factor de riesgo y una enfermedad. Con frecuencia, los ensayos clínicos aleatorios utilizan como criterio de medición de los resultados de una intervención la presencia o ausencia de un episodio adverso o favorable[23]. En el caso de una intervención beneficiosa, su efecto debería reflejarse en la existencia de una menor proporción de episodios en el grupo de intervención que en el grupo de control. Pero la manera más simple

de saber la magnitud de un efecto es calcular la diferencia absoluta entre ambas proporciones, lo que se conoce como reducción absoluta del riesgo (RAR). Otra medida de efecto, comúnmente referida en las publicaciones, es la reducción relativa del riesgo (RRR), que ajusta la RAR a una escala relativa y que expresa la reducción del riesgo respecto al riesgo en el grupo de control.

- *La precisión del estudio. ¿Cuán precisos fueron los resultados?* Todas las medidas que acabamos de ver vienen referidas o pueden calcularse a partir de los resultados contenidos en el artículo que estamos evaluando. Pero a pesar de que son las mejores medidas disponibles, nunca podremos estar absolutamente seguros de que reflejan la verdadera reducción del riesgo, ya que los ensayos clínicos trabajan con una muestra limitada de pacientes y, por tanto, la estimación obtenida sobre el efecto de un determinado tratamiento no deja de ser un valor puntual. Pero interesa conocer cuál será la estimación del efecto del tratamiento en la población general. Por ello, además de conocer la magnitud del efecto, es necesario conocer la precisión de esa estimación a través de los intervalos de confianza.

Es previsible que el verdadero valor se encuentre en las proximidades de esta estimación, y sea tanto más próximo cuanto mayor sea el tamaño de la muestra. El intervalo de confianza nos va a permitir concretar ese grado de proximidad, y es el

intervalo dentro del que se encuentra la verdadera magnitud del efecto con un grado prefijado de seguridad (95%). Un intervalo de confianza del 95% es aquel en el que en el 95% de los casos, o con una seguridad del 95%, el verdadero valor de la magnitud del efecto se encuentra dentro de ese intervalo. Cuanto más estrecho sea el intervalo de confianza, mayor será la precisión con la que se estima el efecto.

El intervalo de confianza contiene información de interpretación más intuitiva que el clásico nivel de significación estadística ($p < 0{,}05$), ya que el lector puede juzgar cuáles son las estimaciones más favorables y desfavorables de la reducción del riesgo (límites superior e inferior del intervalo). Sin embargo, ni los intervalos de confianza ni el grado de significación deben restar al lector atención sobre la significación clínica de los resultados obtenidos.

1.6.3. Sobre la validez externa del estudio. ¿Son aplicables los resultados del ensayo?

La generalidad de los resultados debe considerarse en términos de validez externa, que condiciona su aplicación a nuestro propio contexto de práctica asistencial. A la hora de juzgar si los resultados del estudio son aplicables a nuestros pacientes, es preciso tener en cuenta si nuestro entorno de trabajo se parece al de los pacientes que participan en el ensayo, si cumplen los criterios de inclusión/exclusión del estudio y si

tienen condiciones de gravedad o comorbilidad diferentes que puedan interferir en el resultado[23].

La valoración de la aplicabilidad de los resultados de un ensayo clínico debe incluir: *a)* la reproducibilidad de los resultados; *b)* el análisis de todas las variables clínicamente importantes, y *c)* la relación entre los beneficios y los perjuicios y costes.

- ***Reproducibilidad de los resultados; ¿son aplicables al paciente que nos ocupa?*** Las características ideales que se procuran en un ensayo clínico aleatorio hacen que su diseño tenga una gran validez interna, pero al mismo tiempo limitan su validez externa. Una forma de valorar la aplicabilidad de los resultados de un ensayo al paciente individual es comprobar si sus características se ajustan a las de los pacientes incluidos en el estudio. Es decir, si nuestro paciente hubiera podido formar parte del ensayo clínico, si cumple todos los criterios de inclusión y ninguno de los de exclusión y si sus características son similares a las que se muestran en la tabla de características basales de los pacientes analizados en el ensayo clínico. En definitiva, se trata de comprobar cuán reproducibles son los resultados del estudio en este paciente en particular.

  En ocasiones, los autores del trabajo presentan en los resultados información diferenciada de subgrupos de pacientes en los que las medidas de efecto son más o menos favorables. Este hecho requiere especial precaución cuan-

do los resultados globales son menos favorables. Solo podremos considerar los resultados del análisis de subgrupos si la diferencia del efecto es grande, no atribuible al azar, el análisis estaba bien planificado a priori y sus resultados concuerdan con los de otros estudios similares.

La reproducibilidad de los resultados debe considerar también si la actuación terapéutica analizada es técnicamente aplicable en el medio en el que trabajamos. Esto es especialmente importante cuando se trata de nuevos tratamientos, sobre todo aquellos que requieren recursos o aprendizaje especializados.

- *Análisis de todos los resultados clínicamente importantes; ¿se han considerado todos?* El ensayo clínico aleatorio debe medir los resultados utilizando variables clínicamente importantes. Para poder planificar la aplicabilidad del estudio es preciso valorar la importancia clínica del parámetro empleado en la medición de los resultados. Así, hay parámetros, como la supervivencia, que no ofrecen dudas en cuanto a su repercusión clínica. Sin embargo, existen otros (escalas subjetivas de valoración de síntomas, indicación de hospitalización), cuyo significado clínico es más problemático.

- *Relación beneficios y perjuicios/costes; ¿merecen la pena los beneficios?* Un aspecto importante a la hora de juzgar la significación clínica de un resultado es considerar

también los posibles efectos adversos no deseados ligados a la intervención aplicada. En la decisión final sobre la aplicabilidad de los resultados tendrán que valorarse la magnitud y la importancia tanto de los beneficios como de los perjuicios y costes.

Imaginemos un ensayo clínico aleatorizado sobre una determinada intervención, hecho con rigor, con una validez interna elevada. Los resultados son relevantes, pues consiguen una mejoría clínica y una marcada recuperación funcional, pero los recursos humanos y económicos dedicados sobrepasan cualquier programa de intervención que podamos sugerir en nuestra área. ¿Cómo podemos aplicar esta intervención a nuestra práctica profesional? En nuestro cuerpo de conocimiento debe figurar la idea de que la intervención es posible en otras condiciones de trabajo y que da resultados positivos en el contexto original de aplicación. Sin embargo, en caso de querer aplicarla a nuestro medio, deberemos modificar sustancialmente el protocolo para adaptarla a nuestro contexto, con el peligro de modificar los resultados positivos obtenidos con el protocolo original

Por último, y especialmente a la hora de valorar la validez externa de los ensayos clínicos aleatorizados, hemos de tener en cuenta el impacto de la utilización del nuevo tratamiento. Para ello se utiliza el concepto de NNT, que corresponde al número de pacientes que deben ser tratados para prevenir un episodio desfavorable o lograr un resultado favorable adicional.

1.6.4. Declaración CONSORT

Para mejorar la calidad de la comunicación de los ensayos clínicos, un grupo de científicos y editores desarrolló el CONSORT *statement (Consolidated Standards of Reporting Trials)*. Buena parte de las preguntas que nos haremos para leer críticamente un ensayo clínico están recogidas en la declaración CONSORT[24].

El objetivo de la guía CONSORT es facilitar la lectura crítica y la interpretación de los ensayos clínicos aleatorizados, orientando a los autores, revisores y editores acerca de cómo debe comunicarse este tipo de estudios. Desde su publicación en 1996 y hasta su última revisión de 2001, diversas revistas y grupos editoriales (incluido el grupo de Vancouver) se basan en esta guía[25].

En la publicación de Altman[26] y en Internet (http://www.consort-statement.org) se puede consultar la justificación de los 22 puntos incluidos en la guía, así como ejemplos publicados de lo que se consideran comunicaciones correctas.

1.7. Lectura crítica de artículos de revisión y metaanálisis

Los artículos de revisión y de metaanálisis son importantes formatos de síntesis del conocimiento. Por ello, la lectura crítica de este tipo de artículos es muy importante, pues no será extraño que algunos autores extraigan conclusiones de la síntesis de estudios que no se adecuan a lo que los estudios individuales indicaban.

De hecho, las revisiones clásicas tienen cierto grado de subjetividad en lo que se refiere a inclusión de estudios y valoración de su validez. El carácter cuantitativo de los metaanálisis hace que gocen de una mayor objetividad, aunque la selección a priori/a posteriori de los trabajos que se incluyen puede ser determinante.

A pesar de que la unidad de análisis son trabajos publicados de estudios observacionales y experimentales, se pueden utilizar igualmente los criterios de validez interna, relevancia y utilidad en nuestra práctica clínica.

1.7.1. Sobre la validez interna de la revisión o metaanálisis

Oxman et al.[27] crearon un test que podríamos resumir en seis preguntas, separadas en dos grupos –preguntas de eliminación y de detalle–, y que constituye una herramienta útil que

puede ayudar al lector en la valoración de la validez interna de una revisión sistemática.

Preguntas de eliminación. La respuesta negativa a estas preguntas supondría descartar, de entrada, la utilidad del artículo analizado, ya que estas preguntas constituyen elementos clave del diseño del trabajo sin los cuales no hay ninguna garantía de la validez de los resultados finales[28].

- *¿Se hizo la revisión sobre un tema claramente definido?* Ello significa que han quedado claramente explicados tres aspectos importantes: *a)* la justificación de la revisión, los antecedentes y la necesidad de la revisión están bien explicados; *b)* el objetivo específico de la revisión, o lo que es lo mismo, la expresión concreta de la pregunta de investigación (qué se quiere resolver), y *c)* los componentes básicos de la revisión, que no son otros que los pacientes (criterios de inclusión y exclusión), la intervención (terapéutica, diagnóstica, potencialmente nociva), la comparación (placebo, tratamiento activo, prueba diagnóstica) y el resultado (episodio o enfermedad que tratar, retrasar, prevenir, diagnosticar).

- *¿Se utilizaron criterios adecuados para la selección de artículos?* Los artículos seleccionados tienen que abordar la pregunta objeto de la revisión. Además, es necesario elegir el diseño apropiado de los artículos seleccionados, que depende del tipo de problema que aborde la revisión: *a)*

si lo que se revisa es un tratamiento, se ha de seleccionar artículos de ensayos clínicos aleatorizados, controlados y con análisis por intención de tratar; *b)* en el caso de una cuestión diagnóstica, los artículos seleccionados habrán de incluir la comparación con un patrón oro y la comparación entre los resultados de las dos pruebas debe ser enmascarada e independiente; *c)* si se aborda una cuestión de etiología o efectos nocivos, se han de elegir artículos de casos y controles, y *d)* si la revisión es sobre una cuestión de pronóstico, los artículos deben estudiar a pacientes en un momento similar en el transcurso de su enfermedad, el tiempo de seguimiento debe ser suficientemente prolongado y su verificación, bastante completa.

El control de la validez interna de una revisión también ha de responder a una serie de **preguntas de detalle**[29].

- *¿Se incluyeron todos los estudios relevantes e importantes?* La revisión sistemática ideal sería la que incluyese todos los artículos relevantes que respondiesen a la pregunta primaria. La identificación de todos estos trabajos constituye, sin duda, el reto fundamental al que se enfrenta todo revisor. Este esfuerzo implica la búsqueda en las bases de datos seleccionadas y la búsqueda manual de todos los trabajos, tesis doctorales, resúmenes de congresos, etc. También es importante, como ya se ha comentado, el contacto directo con investigadores y expertos para la obtención de trabajos no publicados. Por todo ello,

debemos buscar en la revisión a estudio las referencias a este esfuerzo y sus resultados. Obviamente, una revisión sistemática o metaanálisis que haya utilizado en su preparación una sola base de datos y en la que no encontremos información sobre el intento de recabar estudios relevantes por otros métodos tendrá mayor probabilidad de haber perdido para su estudio trabajos relevantes, con lo que habrán aumentado sus posibilidades de sesgos y, con ello, habrá disminuido su validez.

- *¿Se evaluaron la validez y la calidad de los estudios incluidos?* El problema que plantean los artículos que componen una revisión sistemática es su validez. La inclusión de un artículo de calidad cuestionable puede afectar al conjunto de la revisión, sobre todo si es un metaanálisis. Así, cuando se incluyen ensayos clínicos de baja calidad (sin aleatorización oculta), se puede sobreestimar la medida del efecto de un metaanálisis, ya que los ensayos de peor calidad tienden a que los resultados derivados de las intervenciones sean más favorables.

  No hay acuerdo sobre cómo evaluar la calidad de los artículos originales en una revisión sistemática. Si bien, para el caso de los ensayos clínicos controlados aleatorizados se usan las recomendaciones de la Colaboración Cochrane, que tiene en cuenta cuatro posibles sesgos como criterios de calidad: *a)* el sesgo de selección, cuando las características iniciales de los participantes en los grupos de

estudio son distintas, que se previene mediante la ocultación de la distribución aleatoria; *b)* el sesgo de ejecución, que se produce cuando hay diferencias sistemáticas en la atención prestada a los participantes del grupo control con respecto al grupo intervención, que se previene mediante el enmascaramiento de los proveedores de la asistencia sanitaria y de los pacientes; *c)* el sesgo de desgaste, que se produce cuando los abandonos son distintos en el grupo control que en el grupo intervención, y que se corrige cuando se puede efectuar un análisis estadístico por tratamiento asignado o por intención de tratar, y *d)* el sesgo de detección, cuando se producen diferencias en la evaluación del desenlace en los dos grupos, que se previene con el enmascaramiento del evaluador.

Sea cual sea el método de evaluación de calidad, los autores deben comparar los resultados de aquellos estudios de calidad alta con los de los estudios de calidad baja, para apuntar posibles diferencias en las medidas de efecto que puedan deberse a los sesgos de los estudios con calidad inferior.

- *¿Son reproducibles las evaluaciones de los estudios?* La selección de estudios, la evaluación de su calidad y la extracción de datos son procesos sometidos a subjetividad. Por tanto, los instrumentos empleados para ello deben ser lo más válidos posibles. En ausencia de criterios objetivos para establecer su validez, se suele recurrir a que

dos evaluadores hagan cada uno el trabajo por su cuenta para después contrastar sus resultados y resolver sus diferencias, ya sea por acuerdo entre ellos o por el concurso de un tercer observador. En ese sentido, la medida del acuerdo entre los dos observadores puede suponer una medida indirecta de la validez.

- *¿Son similares los resultados en los distintos estudios de la revisión?* Las revisiones reúnen estudios que, aunque tengan en común una serie de características pertinentes al objetivo del estudio, difieren en otras muchas, como pueden ser la gravedad o comorbilidad de los pacientes, su ámbito geográfico y social, la intensidad de la intervención en duración, dosis o medicamento concreto y otros factores. Algunas de esas diferencias pueden ser lo suficientemente grandes para hacer poco razonable un metaanálisis. En cualquier caso, siempre que se encuentra una cierta discordancia entre los resultados de los estudios, los autores deben intentar explorar sus causas para averiguar qué diferencias en las características de los pacientes, de la intervención o del desenlace pueden explicar las discrepancias, y así formular hipótesis para futuros estudios.

1.7.2. Relevancia de los resultados de la revisión o metaanálisis

Antes de conocer los resultados globales de la revisión, se deben conocer las características de los estudios que la componen: tipo de diseño, características de los participantes, particularidades de la intervención y de la comparación (dosis, duración), tiempo de seguimiento, definiciones de los desenlaces y cualquier otro dato del estudio que se considere relevante[28]. Una vez hecho esto, y tras estar seguros de la validez del estudio, debemos fijarnos en la *magnitud* o *medida del efecto* y la *precisión* de la asociación observada para decidir sobre la relevancia de los resultados. Y resulta relativamente fácil en las revisiones sistemáticas que incluyen metaanálisis de los estudios individuales, con una estimación global y ponderada. Estas *medidas del efecto* se ponderan según la importancia de cada estudio, cuyo peso suele venir dado por el inverso de la varianza, que tiene que ver con el tamaño de muestra, de modo que aquellos estudios con un tamaño de muestra mayor son los que aportan una contribución mayor en la ponderación final[28]. A diferencia del metaanálisis, en las revisiones narrativas se tiende a clasificar los estudios en dos grupos: los que dan resultados favorables y los que los dan desfavorables, sin ninguna consideración a su importancia relativa.

En la mayoría de las revisiones los resultados se presentan mediante un *análisis de sensibilidad* mediante el cual se analizan los estudios según compartan determinadas caracterís-

ticas de calidad, pacientes, intervención o desenlace. El propósito de estos análisis es explorar posibles diferencias en los resultados que pudieran ser atribuibles a las diferencias en dichos factores.

Los *análisis de subgrupos*, aunque tentadores para conocer los resultados específicos de determinadas características iniciales de los pacientes (nunca factores de riesgo), están sujetos a sesgos y sus resultados muchas veces no tienen más mérito que el de formular hipótesis para futuros estudios. Por tanto, los *análisis de subgrupos post hoc* o a posteriori tienen escaso crédito y pueden ser engañosos.

Si hay sospecha de *sesgo de publicación*, la revisión debería explicar los procedimientos empleados para su evaluación. Asimismo, los resultados han de incluir una descripción de los episodios adversos por grupo, con especial énfasis en los graves.

- ***¿Fueron precisos los resultados?*** Al igual que en los ensayos clínicos, la precisión de los resultados la aportan sus intervalos de confianza, que informan del grado de incertidumbre de los resultados. Cuando los intervalos de confianza son amplios, los resultados son poco precisos. Son varias las razones por las que los resultados de una revisión o metaanálisis son imprecisos. El principal motivo en un metaanálisis es su pobre potencia estadística, que tiene lugar cuando el número de estudios es escaso o el tamaño de la muestra de los estudios metaanalizados

es pequeño. También puede ocurrir cuando el tiempo de seguimiento es insuficiente o cuando la medida de la variable de resultado es imprecisa por sí misma.

1.7.3. Validez externa de la revisión o metaanálisis

Para valorar la utilidad de los resultados en nuestra práctica clínica deberemos tener en cuenta la información recogida en el estudio, como las características de elegibilidad de los pacientes, el ámbito y el lugar donde se recogieron los datos, el tipo de intervención, la definición del desenlace, la representatividad de los casos reclutados, la duración del tratamiento y el tiempo de seguimiento. Todos ellos ayudan a saber si los resultados de la revisión serían aplicables a nuestros pacientes. Cuando ellos o nuestro ámbito no coinciden en lo que respecta a características demográficas, de comorbilidad u otras, la pregunta es: ¿se pueden extender los resultados de esta revisión a nuestra población a pesar de las diferencias? ¿Se pueden extrapolar los resultados de esta revisión a situaciones con mayor tiempo de seguimiento o mayor duración del tratamiento?

Resulta también oportuno comprobar si se tuvieron en cuenta todos los resultados de *importancia clínica*. Estos abarcan desde las variables clínicas importantes que el estudio pretende prevenir, hasta episodios adversos, junto con otros aspectos como la calidad de vida relacionada con la salud, la

satisfacción del paciente y los costes económicos. La cuestión que como lectores hemos de plantearnos es, en caso de que no se hayan tenido en cuenta todos los resultados de importancia clínica, ¿en qué medida afecta la aplicación hipotética de los resultados a nuestra práctica clínica?

Aunque el propósito de una revisión es establecer exhaustivamente las pruebas existentes sobre una determinada cuestión, su interpretación también necesita conocer qué otros estudios previos han abordado esa cuestión u otras directamente relacionadas. En un paciente particular los resultados de una revisión se deben expresar en términos que incluyen su riesgo inicial y los efectos terapéuticos. En los metaanálisis de ensayos clínicos el impacto se puede estimar con medidas como el NNT, aunque deben ser considerados con cautela en función de la homogeneidad de los estudios incluidos.

En el *balance final de beneficios y riesgos*, antes de tomar una decisión se debe tener en cuenta la incorporación a la decisión de las preferencias y los valores del propio paciente. Hay pruebas de que la participación activa del paciente en la toma de decisiones mejora su calidad de vida y sus resultados.

1.7.4. Declaración QUORUM

Dado que los metaanálisis y las revisiones sistemáticas son herramientas de reciente aparición, muchos médicos no se encuentran preparados para la interpretación y apreciación

de los distintos matices a la hora de valorar sus resultados. De ahí la necesidad de diseñar unas directrices de interpretación de las revisiones sistemáticas y los metaanálisis. QUORUM *(Quality of Reporting of Metaanalysis)* nace con esta misión. Incluye una lista de comprobación estructurada con 18 ítems que tanto autores como editores deberían considerar a la hora de publicar un metaanálisis[30]. Sin embargo, la declaración QUORUM no parece haber logrado el mismo grado de aceptación por parte de los editores de revistas biomédicas, aun cuando el número de revisiones y metaanálisis que se publican anualmente es muy elevado y en número creciente. En el momento de su publicación, el grupo de trabajo QUORUM estableció la necesidad de una revisión de y una actualización periódica de la lista de comprobación inicial. Posteriormente, en el año 2009 se publicó la declaración PRISMA *(Preferred Reporting Items for Systematic reviews and MetaAnalyses)*, una actualización y ampliación de QUORUM consistente en una lista de comprobación estructurada de 27 ítems[31]. PRISMA sigue la estructura básica de un artículo científico, detallando los contenidos específicos que deberían comunicarse en cada sección. También propone la utilización de un diagrama de flujo para ilustrar el proceso de la revisión. La finalidad de PRISMA es ayudar a los autores de una revisión sistemática (con o sin metaanálisis) a partir de ensayos clínicos u otro tipo de estudios a recordar los aspectos clave que deberían considerar a la hora de publicar su trabajo en forma de artículo en una revista biomédica, así como también a los editores de

estas revistas para mejorar el proceso de revisión de los manuscritos y el producto final que se publica.

La asunción de PRISMA es doble. Por una parte, asume que las revisiones sistemáticas son una pieza de información científica clave para una toma de decisiones clínicas (prescripción) y sanitarias (asignación de recursos y formulación de políticas de salud) bien fundamentada. Por otra, y en íntima relación con lo anterior, que las revisiones sistemáticas son un modelo de investigación formal de especial utilidad, dada su eficiencia para obtener respuestas válidas y reducir el grado de incertidumbre acerca de los efectos de los tratamientos e intervenciones en el menor tiempo posible. Por ello, resulta esencial garantizar la calidad y rigor de los métodos empleados, del mismo modo que lo exigimos para interpretar adecuadamente un estudio de investigación primaria. De ahí la importancia de asegurar un trabajo completo y detallado en el que se expliquen con suficiente claridad los métodos y procedimientos empleados a lo largo del proceso de una revisión, comenzando por la búsqueda bibliográfica y terminando por el análisis e interpretación de los resultados. Además, estos deberán ser presentados de forma clara, concisa e informativa para quienes después utilizarán la revisión para la toma de decisiones[32].

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(2ª parte)

2

BIBLIOGRAFÍA

2. BIBLIOGRAFÍA

1. Walker AM. Observation and inference. An introduction to the methods of epidemiology. Chestnut Hill, MA: Epidemiology Resources Inc.; 1991.
2. Hernández-Ávila F, Garrido-Latorre F, López-Moreno S. Diseño de estudios epidemiológicos. Rev Salud Publ Mex. 2000;42:144-55.
3. Ciapponi A. Guía de lectura crítica de estudios observacionales en epidemiología (primera parte). Evid Actual Pract Ambul. 2010;13:135-40.
4. Miettinnen OS. Theoretical epidemiology. Principles of occurrence research medicine. Nueva York: A Wiley Medical Publication; 1985.
5. Fontcuberta J, García-González AL. Lectura crítica de otros tipos de estudios: estudios observacionales. Revisiones narrativas. Paneles de expertos. Documentos de consenso (cap. 14; pp. 389-412). En: Atención sanitaria basada en la evidencia. Su aplicación a la práctica clínica. Grupo de Atención Sanitaria Basada en la Evidencia. CSRM; 2006.
6. Etapas del desarrollo de una vacuna. Disponible en URL: http://www.elergonomista.com/saludpublica/etapas.htm
7. Martín, J.L.R., Tobías, A. y Seoane, T. (Coords.). Revisiones sistemáticas en ciencias de la vida. Toledo: FISCAM; 2006.
8. Goodman SN. Have you ever Meta-analysis you didn't like? Ann Int Med. 1991;114:244-46.

9. Littell, J.H., Corcoran, J. y Pillai, V. Systematic reviews and meta-analysis. Oxford, UK: Oxford University Press; 2008.
10. Marín Martínez F. El metaanálisis en el ámbito de las ciencias de la salud. Fisioterapia. 2009;31:107-14.
11. Von Elm E, Altman DG, Egger M, et al., en nombre de la Iniciativa STROBE. Declaración de la Iniciativa STROBE (Strengthening the Reporting of Observational Studies in Epidemiology): directrices para la comunicación de estudios observacionales. Gac Sanit. 2008;22:144-50.
12. Fernández E, García AM. Búsqueda y lectura crítica de artículos científicos. Barcelona: Signo; 2006.
13. Lazcano E, Fernández E, Salazar E, et al. Estudios de cohorte. Metodología, sesgos y aplicación. Rev Salud Publ Mex. 2000;42:230-41.
14. MacMahon B, Trichopoulos D. Epidemiology. Principles and methods. 2nd Ed. Boston: Little Brown and Company;1996.
15. Howe GR. Use of computerized record linkage in cohort studies. Epidemiol Rev. 1998;20:112-21.
16. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. Br Med J. 2004;328:39-41.
17. Black N. Why we need observational studies to evaluate the effectiveness of health care. Br Med J. 1996;312:1215-8.
18. Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Epidemiol. 2007;18:805-35.
19. Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. PLoS Medicine. 2007;4:e297. doi: 10.1371/journal.pmed.0040297.
20. Bartlett C, Sterne J, Egger M. What is newsworthy? Longitudinal study of the reporting of medical research in two British newspapers. Br Med J. 2002;325:81-4.
21. Fernández E, García AM. Cómo leer artículos científicos. Ud 8. En: Fernández E., García AM. Ed. Preparación de publicaciones biomédicas. Disponible en: http://www.metodo.uab.es

22. Tello C. Lectura crítica de un ensayo clínico. (Cap. 10; pp. 275-96). En: Atención sanitaria basada en la evidencia. Su aplicación a la práctica clínica. Grupo de Atención Sanitaria Basada en la Evidencia. CSRM; 2006.
23. Ochoa-Sangrador C, Martín-Muñoz P, Ruiz-Canela J. Lectura crítica de artículos. Rev Pediatr Aten Primaria. 2007;9 Supl 2:S151-60.
24. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. The CONSORT Group. Lancet. 2001;357:1191-4.
25. Cobos-Carbó A. Ensayos clínicos aleatorizados (CONSORT). Med Clin (Barc). 2005;125(Supl. 1):21-7.
26. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134:663-94.
27. Oxman AD, Cook DJ, Guyatt GH. Users' Guides to the Medical Literature: VI. How to Use an Overview. JAMA. 1994; 272:1367-71.
28. García-López F, Quereda C. Apuntes para la lectura crítica de una revisión sistemática. Nefrología. 2007;27:417-24.
29. Fuensanta G. Lectura crítica de un metaanálisis y de una revisión sistemática. (Cap. 11; pp. 297-328). En: Atención sanitaria basada en la evidencia. Su aplicación a la práctica clínica. Grupo de Atención Sanitaria Basada en la Evidencia. CSRM; 2006.
30. Moher D, Cook DJ, Eastwood S, et al. Improving the quality of reports of metaanalyses of randomised controlled trials: the QUORUM statement. Lancet. 1999;354:1896-900.
31. PRISMA. Transparent Reporting of systematic Reviews and MetaAnalises. Disponible en: http://www.prismastatement.org/index.htm
32. Urrútia G, Bonfill X. La declaración PRISMA: un paso adelante en la mejora de las publicaciones de la Revista Española de Salud Pública. Rev Esp Salud Pública. 2013;87(2):99-102 E.

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(2ª parte)

3

AUTOR

José Luis Mateos

Nacido en Madrid en 1966, casado y con una hija.

Experto en la preparación de conferencias y material documental para personal sanitario, cuenta en su haber con más de una treintena de artículos publicados en revistas americanas y europeas de diversa índole. Es doctor en biología por la Universidad Complutense de Madrid, investigador asociado del Instituto de Bioquímica y Biología Molecular de la Universidad Libre de Bruselas y Master Science in Biomedicine por la Universidad de Newcastle. Por su trabajo actual como director clínico de una *biotec* de origen alemán en España, su perfil está plenamente dedicado al apasionante y novedoso campo de la terapia celular de aplicación en cáncer hematológico y de órgano sólido y en medicina regenerativa.

Lectura crítica de un artículo científico

¿Cómo analizar un artículo publicado?

(2ª parte)

4

NOTAS

NOTAS

NOTAS

NOTAS